

One Hundred Years of Schizophrenia: A Meta-Analysis of the Outcome Literature

James D. Hegarty, M.D., M.P.H., Ross J. Baldessarini, M.D.,
Mauricio Tohen, M.D., Dr.P.H., Christine Wateraux, Ph.D., and Godehard Oepen, M.D.

***Objective:** This study was undertaken to assess the twentieth-century literature on outcome in schizophrenia for historical trends that might be associated with changes in diagnostic and therapeutic practice and to test the hypothesis that both improved biological treatment and changes in diagnostic criteria have influenced outcome. **Method:** Meta-analysis of the international literature on outcome in schizophrenia or dementia praecox from 1895 to 1992 identified 821 studies; 320 of these, with 51,800 subjects in 368 cohorts, met the inclusion criteria for the study. **Results:** Only 40.2% of patients were considered improved after follow-ups averaging 5.6 years (range=1-40). Outcome was significantly better when patients were diagnosed according to systems with broad criteria (46.5% were improved) or undefined criteria (41.0% were improved) rather than narrow criteria (27.3% were improved). The proportion of patients who improved increased significantly after mid-century (for 1956-1985 versus 1895-1955, 48.5% versus 35.4%), probably reflecting improved treatment as well as a broadened concept of schizophrenia. However, in the past decade, the average rate of favorable outcome has declined to 36.4%, perhaps reflecting the re-emergence of narrow diagnostic concepts. **Conclusions:** Overall, less than half of patients diagnosed with schizophrenia have shown substantial clinical improvement after follow-up averaging nearly 6 years. Despite considerable gains in improvement rates after mid-century, there has been a decline since the 1970s. These historical changes probably reflect improved treatment, shifts in diagnostic criteria, and selection bias related to changes in health care.*

(Am J Psychiatry 1994; 151:1409-1416)

Presented in part at the 145th annual meeting of the American Psychiatric Association, Washington, D.C., May 2-7, 1992. Received Nov. 3, 1992; revisions received Aug. 18, 1993, and Jan. 25, 1994; accepted Feb. 7, 1994. From the Consolidated Department of Psychiatry and the Neuroscience Program, Harvard Medical School, Boston; the Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston; and the Laboratories for Psychiatric Research and the Bipolar and Psychotic Disorders Program, Mailman Research Center, McLean Division of the Massachusetts General Hospital. Address reprint requests to Dr. Hegarty, McLean Hospital, 115 Mill St., Belmont, MA 02178.

The authors thank Dr. Franca Centorrino and Dr. Elizabeth M. Sajdel-Sulkowska for assistance with translation of non-English documents and Thelma Charen, M.A., Lyn Dietrich, M.S.L.S., Debbie Penningo, Mila Cason, M.A., and Jeanmarie O'Donnell, B.S., for bibliographic assistance.

Supported in part by NIMH grants MH-31154, MH-47370, and MH-48444 and by a Young Investigator Award from the National Alliance for Research on Schizophrenia and Depression.

Clinical course and outcome in schizophrenia (formerly dementia praecox) have been the subject of extensive investigation through this century. Kraepelin (1) considered dementia praecox to be a chronic or progressive illness leading to severe impairments in cognitive and social functioning, with findings of clinical improvement during follow-up in not more than 17% of cases. While the progress or course of the illness varied, a poor outcome was considered to be almost inevitable eventually or was inherent in the diagnostic conceptualization. For example, in a large early survey of 395 hospitalized patients in 1912, Stearns (2) lamented the "apparent hopelessness of the disease dementia praecox." Early treatments, such as injection of adrenaline (3), inducing fever (4), and vasectomy (5),

proved wholly ineffective, and pessimistic conclusions dominated the prognostic literature for decades. In the 1930s the introduction of insulin coma (6), chemoconvulsive therapy (7), and ECT (8) provided the first effective treatments for psychotic disorders. These more effective interventions, as well as the growth of a broader conceptualization of schizophrenia introduced by Eugen Bleuler in 1911 (9), were widely adopted, and the prognosis for schizophrenia began to appear less bleak by the late 1930s (10). The development of specific antipsychotic medicines, beginning with reserpine and chlorpromazine in the 1950s (11–15), introduced a new era in treatment, and hopes soared for improved outcomes as well as short-term benefits in acute phases of psychotic illness (16–18).

During the 1970s two large, long-term outcome studies (19, 20) found significant clinical improvement in nearly 50% of schizophrenic patients after 10 years. These reports added to a growing literature supporting an improved prognosis for patients with schizophrenia in the contemporary era (21–25). Credit was given to improved treatment as well as deinstitutionalization and new clinical appreciation of psychosocial factors in chronic psychotic disorders, including family relations and expressed emotion (26). It may also be important that most of these studies (19–26) used relatively broad or cross-sectional diagnostic criteria, in contrast to the most recent American diagnostic methods (DSM-III and DSM-III-R), which attend to duration of symptoms and course of illness in what might be termed a "neo-Kraepelinian" trend. Therefore, it is of interest to address historical trends in studies of outcome in schizophrenia and to attempt to ascertain the impact of diagnostic and treatment methods.

Reviews of some of the literature on outcome in schizophrenia have been carried out by others (27–32), but to date there have been no studies incorporating the available international outcome data in a meta-analysis with explicit inclusion and exclusion criteria. Nor has the possible impact of evolving diagnostic concepts been incorporated into such analyses. Accordingly, we undertook the current study, applying recommended contemporary methods of meta-analysis for clinical studies (33) and considering available information on diagnosis and treatment in particular.

METHOD

We undertook a comprehensive review of the twentieth-century literature regarding outcome in schizophrenia. Studies published before 1966 were identified by a title search of *Index Medicus* under the headings Dementia Praecox and Schizophrenia, with the subheadings outcome, course, and prognosis. Historically, literature on schizophrenia was indexed under the subject heading Dementia Praecox for articles listed in the *Current List of Medical Literature* from 1950 to 1953 and under Schizophrenia for articles listed in the *Current List of Medical Literature* from 1954 to 1959 and in *Index Medicus* from 1960 to the present (34–38). Studies published since 1966 were identified through a MEDLINE computerized literature search in which the same title search criteria were used. The entire search encompassed the years 1895 through 1991 and yielded 686 primary reports;

additional references (N=135; 16.5%) found in primary reports and meeting the title search criteria or identified as outcome studies also were examined, to provide a total of 821 reports for preliminary consideration.

Inclusion Criteria

Studies identified by the title search were included for analysis if they met the following criteria. 1) They included only patients with diagnoses of schizophrenia or dementia praecox or gave outcome data specific to these diagnostic groups; studies mixing schizophrenic patients with patients who had manic-depressive, schizoaffective, or other psychoses were excluded. 2) The mean length of follow-up was at least 1 year (i.e., cross-sectional or census studies were excluded). 3) Less than 33% of the subjects were lost to follow-up, and there were 15 or more patients at completion of follow-up. 4) Subjects were not selected a priori for good or poor outcome. 5) Explicit numerical data on clinical, social, or vocational outcome were available, so that patients could be assigned to discrete outcome groups (i.e., studies that used correlative relationships or the category "discharged from hospital and not readmitted" were excluded). 6) Both experimental and naturalistic studies had to specify the treatment used throughout the follow-up.

Of the total of 821 studies preliminarily identified by the search criteria, 260 were excluded because they were not outcome studies and eight reports could not be located; the 553 remaining reports were reviewed in detail. The exclusion rate for studies identified by the *Index Medicus* search (1895–1965) was lower (39.5%) than the rate for later studies (1966–1992) found by the computerized MEDLINE search (69.0%), with a moderate correlation between exclusion rate and decade of study ($r=0.56$, $df=8$, $n.s.$). Most studies excluded after 1965 were eliminated because the follow-up was less than 1 year.

Overall, 320 studies (57.9%) met all of the inclusion criteria for analysis; they represent 51,800 subjects from 368 cohorts and 311,400 person-years of follow-up. Non-English-language reports of studies (N=85) that were translated for this analysis included 42 in German, 12 in French, nine in Italian, seven in Danish, five in Norwegian, three in Polish, three in Russian, three in Spanish, and one in Finnish. In addition, 70 reports on non-English-speaking cohorts were published in English. The translated studies that met the inclusion criteria (N=37) represent 11.6% of the 320-study database. (A complete 327-reference bibliography is available from the first author on request.)

Definition of Outcome

The difficulty of accurately and consistently defining clinical outcome in schizophrenia has been reviewed (39). Given the varied criteria used to express outcome in the hundreds of studies we considered, a generalized standard was required. Our determination of outcome made use of clinical, social, or vocational criteria as reported in the studies analyzed. By consensus, we translated each study's outcome data to ratio measures of improvement (the number of patients considered improved divided by the total number per cohort), using criteria that suggested attainment of substantial levels of functioning and freedom from psychotic symptoms. Thus, patients considered as "improved" in follow-up had to have been described as recovered, in remission, well without residual symptoms, minimally or mildly symptomatic, improved without significant deficit, socially recovered, or working or living independently.

Diagnostic Systems

If specific diagnostic criteria were defined, the studies were assigned to the schizophrenia diagnostic systems of the following: Kraepelin (dementia praecox) (1), Langfeldt (40), Bleuler (9), Schneider (first-rank symptoms) (41), Leonhard (42), Mayer-Gross (43), ICD-8, ICD-9, DSM-II, DSM-III, DSM-III-R, the Present State Examination (PSE) (44), the Feighner criteria (45), and the Research Diagnostic Criteria (RDC) (46). These diagnostic systems were then

assigned to either the so-called narrow category (Kraepelinian, with some implication for long-term course of illness) or broad category (non-Kraepelinian) (47). Diagnostic systems requiring at least 6 months of illness before diagnosis of schizophrenia were considered Kraepelinian and include those of Kraepelin, Langfeldt, and Feighner, as well as DSM-III and DSM-III-R. Non-Kraepelinian systems were defined as more symptom-based, cross-sectional diagnostic approaches with a brief duration requirement (≤ 1 month) or no requirement for duration of prodromal or active psychotic illness. These include Bleuler's schizophrenia, Leonhard's systematic schizophrenia, Schneider's first-rank criteria, Mayer-Gross's schizophrenia, ICD-8, ICD-9, DSM-II, the PSE/CATEGO system, and the RDC. Studies that did not use clearly defined diagnostic systems were categorized as having unspecified criteria.

Stratification by Decade, Length of Follow-Up, and Treatment

Only cohorts with ≤ 10 years of follow-up (314 cohorts, or 85.3% of the total of 368) were used for inter-decade comparisons. The mean year of follow-up was calculated as the average of the first and last years of the follow-up period, and decades were defined by their midpoint (e.g., decade 1950 included mean follow-up years falling between 1946 and 1955). Stratification for duration of follow-up included 1 to < 2 , 2-4, 5-9, 10-19, and ≥ 20 years. On the basis of treatment used throughout the period of follow-up, four primary treatment groups were defined: 1) nonspecific (placebo trials, psychotherapy, hydrotherapy, fever therapy, and nonneurological surgery), 2) convulsive (electrically or chemically induced seizures or insulin shock), 3) prefrontal lobotomy, and 4) neuroleptic or atypical antipsychotic drugs.

Statistical Analysis

Data are presented as unweighted means and standard deviations unless stated otherwise. For multiple comparisons, means were compared by Duncan's multiple range test (48). A weighted multiple regression model for predicting percentage of patients improved included the following variables: diagnostic system, type of treatment, and length of follow-up. Weights based on number of subjects, with an estimate of variance, were computed for each study by the method of DerSimonian and Laird (49), and overall variance is given as the standard error of the mean (SE). To ensure homogeneity of variance in parametric tests, data were subjected to an arc sine variance-stabilizing transformation (50). Use of standard analysis of variance (ANOVA) and post hoc tests is cited in the Results section. All *p* values are based on two-tailed tests, with nonsignificance defined as ($p \geq 0.05$).

RESULTS

Table 1 summarizes characteristics of the outcome studies included in this analysis. The number of cohorts of psychotic patients that met the study selection criteria per decade (excluding the incomplete current decade, 1985-1995), averaged 38.7 ($SD=27.1$), with a maximum of 94 during the 1930s, following the introduction of insulin coma and ECT. The average number of subjects per cohort studied declined significantly over the century, from 250 in 1895-1915 to 56 since 1986 (by decade, $r=-0.98$, $N=368$, $p<0.0001$). The diagnosis of schizophrenia or dementia praecox in this century was made without explicit diagnostic criteria (unspecified) in 50.0% of the studies meeting the selection criteria. Kraepelinian diagnostic systems were applied in 21.7% of the studies, and specified non-

TABLE 1. Characteristics of the Schizophrenia Outcome Studies Analyzed

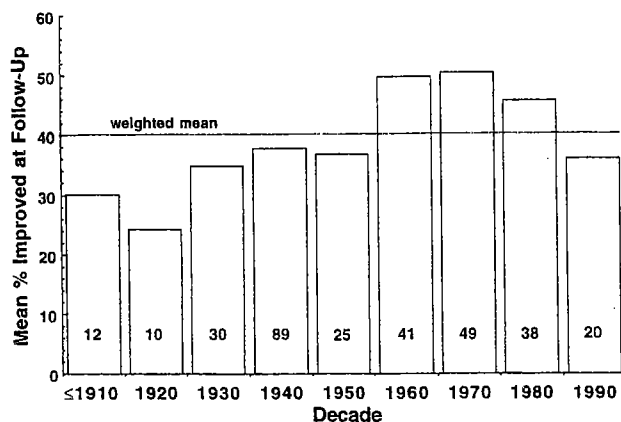
Item	Cohorts (N=368)	
	N	%
Decade in which studies were completed		
1895-1905	6	1.6
1906-1915	6	1.6
1916-1925	13	3.5
1926-1935	37	10.1
1936-1945	94	25.5
1946-1955	36	9.8
1956-1965	51	13.9
1966-1975	63	17.1
1976-1985	42	11.4
1986-1992	20	5.4
Study location ^a		
United States	137	42.8
United Kingdom	67	20.9
Scandinavia	44	13.7
Germany	24	7.5
Switzerland	11	3.4
Japan	8	2.5
China	5	1.6
Canada	5	1.6
India	4	1.2
Other	15	4.7
Diagnostic system used		
Kraepelinian	80	21.7
DSM-III, DSM-III-R	20	25.0
Feighner	9	11.2
Kraepelin	41	51.2
Langfeldt	10	12.5
Non-Kraepelinian	104	28.3
Bleuler	18	17.3
DSM-II	9	8.7
ICD-8, ICD-9	16	15.4
Leonhard	3	2.9
Mayer-Gross	5	4.8
Present State Examination	33	31.7
Research Diagnostic Criteria	10	9.6
Schneider	10	9.6
Unspecified	184	50.0
Years of follow-up		
1 to < 2	98	26.6
2-4	112	30.4
5-9	89	24.2
10-19	48	13.0
≥ 20	21	5.7
Treatment		
Neuroleptic	173	47.0
Convulsive	101	27.4
Lobotomy	13	3.5
Nonspecific	81	22.0

^aThe numbers for study location are of studies, not cohorts.

Kraepelinian methods appeared in 28.3%. The duration of follow-up ranged from 1 to 40 years (median=3.0 years) and averaged 5.6 years ($SD=6.5$). Most of the studies involved relatively short-term follow-up (< 5 years in 57.0%). Treatment predominantly involved neuroleptics, followed by convulsive and nonspecific methods and lobotomy.

The mean proportion of patients showing a favorable outcome, independent of length of follow-up, across the entire century was 40.2% ($SD=17.6\%$) ($N=368$ cohorts). A decade-by-decade summary (figure 1) supports the further impression that outcome shifted over

FIGURE 1. Mean Percentages of Schizophrenic Patient Cohorts Considered Improved in Follow-Ups of ≤ 10 Years, by Decades of Studies Reviewed^a



^aHorizontal line indicates weighted overall mean for the century (40.1%; $N=314$ cohorts). Numbers inside bars indicate number of cohorts for each decade. Decades are defined by the midpoint of the decade (e.g., the 1950 decade is 1946–1955). The year of a study was defined as the mean of the first and last years of follow-up (≤ 1910 includes all studies from 1895 through 1915). There was significant variation in outcome by decade ($F=5.9$, $df=8$, 305 , $p<0.0001$).

time. Prior to 1920, studies consistently involved patients diagnosed as having Kraepelin's dementia praecox, for whom no specific treatment was available. Accordingly, between 1895 and 1925, the weighted average proportion of patients who were considered improved at the end of follow-up was only 27.6% ($SE=3.3\%$). In the 1930 decade, when convulsive therapies were introduced, mean outcomes tended to improve to 34.9% ($SE=3.1\%$). This trend continued after mid-century, with a peak mean in the period 1956–1985 (48.5%, $SE=1.7\%$), which was significantly different from the mean for 1895–1955 (35.4%, $SE=1.3\%$) ($t=5.7$, $df=292$, $p<0.0001$), perhaps reflecting the introduction of modern antipsychotic drug therapy (51, 52). However, since 1986 the mean likelihood of a favorable outcome has diminished to only 36.4% ($SE=4.4\%$), or a level that is statistically indistinguishable from that found in the first half of the century ($t=0.2$, $df=146$, n.s.).

Duration of follow-up also was examined as a predictor of clinical outcome. The unweighted likelihood of a favorable outcome tended to decline with longer follow-up, but only slightly and nonsignificantly: 43.7% ($SD=21.3\%$) improved at 1 to <2 years, 40.8% ($SD=19.2\%$) at 2–4 years, 38.3% ($SD=16.4\%$) at 5–9 years, 37.4% ($SD=18.6\%$) at 10–19 years, and 37.3% ($SD=13.6\%$) at ≥ 20 years ($F=1.5$, $df=4$, 363 , n.s.).

More striking differences were found after stratification for treatment group and diagnostic system (table 2). Unweighted analysis found neuroleptic treatment to be associated with the highest mean percentage of improved subjects (45.6%), followed by convulsive and nonspecific treatments and prefrontal lobotomy. A weighted ANOVA found a highly significant effect of

diagnostic system and the impact of all treatments combined, with no significant interaction between treatment and diagnostic system. Post hoc analysis indicated that the combined results of neuroleptic and convulsive treatments were better than those of lobotomy and non-specific methods.

Studies using broad non-Kraepelinian criteria consistently showed better outcomes than those using narrow Kraepelinian criteria (ratio=1.70; $t=8.3$, $df=178$, $p<0.0001$) (table 2). This difference held throughout the century—from the early era before effective somatic therapies (up to and including 1925), through the introduction of convulsive therapies (1926–1955), through the introduction of neuroleptics (1956–1975), to the contemporary era (after 1975)—and was highly significant by ANOVA (figure 2).

Finally, a weighted multiple regression model was applied to examine diagnostic system, duration of follow-up, and neuroleptic treatment as predictors of improvement. Diagnostic system was a significant predictor of outcome ($F=24.4$, $df=2$, 360 , $p<0.00001$), as was neuroleptic treatment ($F=31.3$, $df=1$, 360 , $p<0.00001$), but duration of follow-up was not ($F=2.0$, $df=4$, 360 , $p=0.09$).

DISCUSSION

The extensive data we analyzed support several general impressions regarding clinical outcome in schizophrenia during the twentieth century. The main impression is that less than one-half of patients diagnosed with this disorder have shown substantial clinical improvement after follow-up averaging nearly 6 years (table 2). Although there were substantial gains in favorable outcome after the 1920s, rates of improvement declined considerably after the 1970s (figure 1). These biphasic trends may reflect the impact of more effective treatments and historical changes in diagnostic criteria during the past century (table 2 and figure 2), as well as possible effects of sampling bias.

Between the 1920s and the 1970s, favorable outcome in schizophrenia (patients rated clinically improved) doubled, from an average of 24.3% to 50.5% of cases (figure 1). Such gains have been attributed to new treatments (17, 18), particularly neuroleptics since the 1950s (51, 52), and an increased emphasis on family and community interventions since the 1960s (26). In the late 1970s Manfred Bleuler (19) noted the apparent disappearance of the most severe, or catastrophic, forms of schizophrenia and proposed that modern treatment might be responsible. It has been hypothesized that early and consistent treatment with antipsychotic medications may favorably alter the natural course of schizophrenia (51), and a recent analysis by Wyatt (52) of 22 follow-up studies involving first-episode patients (21 of which applied broad diagnostic criteria) found that early intervention with neuroleptics increased the likelihood of a favorable long-term course in such patients. The poor outcome long associated with insidi-

